Big data in drug discovery & development

David Wild, Ph.D.
Associate Professor, Indiana University School of Informatics & Computing
Email djwild@indiana.edu; Website djwild.info; Twitter @davidjohnwild

Director, Integrative Data Science Lab
Director Data Science Academic Programs
CEO, Data2Discovery Inc

ids1.soic.indiana.edu
datascience.soic.indiana.edu
d2discovery.com
About me

Associate Professor of Informatics & Computing at Indiana University
Director of Data Science Academic Programs
Director, Integrative Data Science Laboratory
CEO, Data2Discovery Inc.

Background in cheminformatics, pharmaceutical research, linked data, semantic technologies and data science

http://djwild.info
IU Integrative Data Science Lab

Integrative Data Science brings together
• heterogeneous datasets
• expertise from different disciplines
• novel data science tools and technologies
to solve real world problems.

Current research focus areas include:
**Precision Drug Intervention** and
**Smart Cities, Health and Emergency Response**

idsl.soic.indiana.edu
Outline

• Overview of drug discovery & development
  • And why this is the hardest data science problem ever!
• Data sources – chemistry and biology (“preclinical”)
  • Chemistry is weird (cheminformatics)
  • Biology is weird (bioinformatics)
• Connecting the dots
  • How chemical and biological entities can be mapped together
• Beyond preclinical
  • Patients are people too
Learning goals for this section

• Know the four “big picture” epochs in disease treatment
• Know the major stages in how drug discovery is currently done in pharmaceutical companies
• Understand some of the radical changes happening in healthcare and how they open up opportunities for data science
• Preclinical data: have basic knowledge of special representations of chemistry and biology data, and how different entities can be mapped together
• Patient data: understand the basic kinds of patient / people data and how they differ from preclinical sources
Epochs in disease treatments

- **Empirical** – up until 1960’s
  - 754 First pharmacy opened in Baghdad
  - Late 1800’s – major pharmaceutical companies, mass production
  - 1900-1960 – major discoveries (insulin, penicillin, the pill …)

- **Rational** – 1960’s to 1990’s
  - “Lock and key” strategy:
    - Target protein ➔ Lead compound ➔ Pharmacology ➔ Humans
    - Computational Drug Discovery
    - Biggest success HIV (RT, protease inhibitors)

- **Big Experiment** – 1990’s to 2000’s (rational model at scale)
  - High throughput screening
  - Microarray Assays
  - Gene Sequencing and Human Genome Project (now NGS)

- **Integrative** – 2010’s onwards
  - Not just locks and keys – everything is connected
  - Using all data from molecule to patient
  - Revolves around the patient - you
Rational Drug Discovery & Development

Identify disease

Isolate protein involved in disease (2-5 years)

Find a drug effective against disease protein (2-5 years)

Preclinical testing (1-3 years)

Formulation & Scale-up

File IND

Human clinical trials (2-10 years)

File NDA

FDA approval (2-3 years)
Rational Drug Discovery & Development

**DISCOVERY**
- Identify disease
- Isolate protein involved in disease (2-5 years)

**DEVELOPMENT**
- Preclinical testing (1-3 years)
- Find a drug effective against disease protein (2-5 years)
- Formulation & Scale-up

**CLINICAL TRIALS**
- Human clinical trials (2-10 years)
- FDA approval (2-3 years)

**MARKETING >>**
Current healthcare model (not data driven)

Pharmaceutical Research
- Discovery
- Clinical Trials
- Market

Artificial Interventions

Healthcare Delivery
- Providers, Payors, Hospitals, Doctors, Patients

Diagnoses and Treatment Plans

Public Health
- Environment, society, policy, economics

Factors affecting health of populations

Study of people
- Study of chemistry & biology
Emerging healthcare model

Pharmaceutical Research
- Phenotypic Drug Discovery
- Adverse Event Prediction
- Population Health
- Quantified Self

Healthcare Delivery
- Combination therapy
- Risk Adjustment
- Performance-based Pricing
- Outcome Analytics
- Nudge Economics

Public Health

How can I stay healthy & get better when I do get sick?
What does this mean for data scientists?

• Current model is siloed and relies on deep siloed domain knowledge, particular with molecular data
• Emerging need for people cross trained / agile in using molecular, patient, and other data sources
• Current opportunities are siloed by organization (pharmaceutical company, healthcare provider, health insurance, academic researchers etc) but the walls are breaking down rapidly.
• “We’re not a health insurance company, we are a data company” [health insurance company VP]
• “We’re not a health insurance company, we are a wellness company” [health insurance chief medical officer]
Preclinical data

• Representing chemical information
  • Cheminformatics
  • 2D and 3D chemical structures

• Representing biological information
  • Bioinformatics
  • Proteins, genes, pathways
Learn cheminformatics!

Cheminformatics is the study of all aspects of the representation and use of chemical and related biological information on computer. It has applications in drug discovery, health, data mining and many other areas.

This site gives you links to some resources to get you started in learning about cheminformatics - a free online course, a low-cost eBook, an established open access journal and a Google community for discussion about educational options. Enjoy learning about this rapidly developing field!
Historical ways of representing chemicals

- **Trivial name**, e.g. Baking Soda, Aspirin, Citric Acid, etc. Identifies the compound, but gives no (or little) information about what it consists of.

- **Chemical formula**, e.g. $\text{C}_6\text{H}_{12}\text{O}_6$. Specifies the type and quantity of the atoms in the compound, but not its structure (i.e. how the atoms are connected by bonds).

- **Systematic name**, e.g. 1,2-dibromo-3-chloropropane. Identifies the atoms present and how they are connected by bonds.
2D structure diagram

Trivial name:
- tyrosine

Systematic names:
- $\beta$-($p$-hydroxyphenyl)alanine
- $\alpha$-amino-$p$-hydroxyhydrocinnamic acid
Linear notations

- Represent the atoms, bonds and connectivity of a molecule in a linear text string
- Consise representation
- Originally designed for manual command line entry into text-only systems
- Now an excellent format for file and database storage (e.g. can be held in a spreadsheet cell, on one line of a text file, or in an Oracle database text field)
SMILES

(one possible) SMILES for this structure is
OC (=O) C (N) CC1=CC=C (O) C=C1

Can identify any chemical structure

There can be several ways of writing the same structure in SMILES (although a system of generating canonical SMILES) exists
SMILES Homepage

http://www.daylight.com/dayhtml/smiles/index.html

• Official Syntax Guide
• Tutorial
• Examples
• Resources
Internal computer representations

• Atom Lookup Table
  • Stores the atomic names (and possibly other info such as valence, charge, etc.) of all atoms in the molecule

• Connection table
  • Indicates which atoms are connected to which other atoms, and with what kind of bond
## Atom Lookup Table

![Diagram of a molecule with labeled atoms]

<table>
<thead>
<tr>
<th>Atom</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td>2</td>
<td>C</td>
</tr>
<tr>
<td>3</td>
<td>C</td>
</tr>
<tr>
<td>4</td>
<td>N</td>
</tr>
<tr>
<td>5</td>
<td>C</td>
</tr>
<tr>
<td>6</td>
<td>O</td>
</tr>
<tr>
<td>7</td>
<td>C</td>
</tr>
<tr>
<td>8</td>
<td>C</td>
</tr>
<tr>
<td>9</td>
<td>C</td>
</tr>
<tr>
<td>10</td>
<td>C</td>
</tr>
<tr>
<td>11</td>
<td>O</td>
</tr>
</tbody>
</table>

Note that Hydrogens are not normally stored explicitly. Their presence is inferred from the valence of the atom.
### Connection table

![Chemical Structure](image)

<table>
<thead>
<tr>
<th>Atom</th>
<th>Label</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>C</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>C</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>N</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>O</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>C</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>C</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>O</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Topological Graph Theory

- Branch of mathematics particularly applicable to the sciences and computer science
- Study of “graphs” which consist of a set of “nodes” and a set of “edges” joining pairs of nodes
Structure Diagrams as Graphs

- 2D structure diagrams very like topological graphs
  - atoms ↔ nodes
  - bonds ↔ edges
- Terminal hydrogen atoms are not normally shown as separate nodes (“implicit” hydrogens)
  - reduces number of nodes by ~50%
  - “hydrogen count” information used to colour neighbouring “heavy atom” atom
- separate nodes sometimes used for “special” hydrogens
  - deuterium, tritium
  - hydrogen bonded to more than one other atom
  - hydrogens attached to stereocentres
Characterising 2D Structures with Fingerprints

- A “fingerprint” is made up of a set of descriptors for a molecule. Each descriptor describes (usually the presence or absence of) a particular 2D structural feature.

- Most fingerprints are binary strings made up of zeros and ones

- The fingerprint characterises the molecule, but doesn’t uniquely describe it. It is useful in many applications we will come to later, e.g. similarity, clustering, diversity.
Simple fingerprints (Structural Keys)

- A pre-defined dictionary of fragments is used, and each bit in a bitstring represents the presence or absence of that fragment in a particular compound.
Types of 2D searching

• Structure search
  • “Is this structure in the database?”

• Substructure search
  • “Find me all of the structures that contain this substructure”

• Similarity search
  • “Find me all of the structures that are similar to this one”
Structure search

- Looking for a particular structure in a database
  - Searching proprietary databases or commercial databases

- e.g. *is this structure in the database?*

- Mathematically, the connection table can be considered a *graph*, and this is a *graph isomorphism* problem (solved)
Substructure search

• Looking for all structures that contain one or more particular structural fragments

  • e.g. which structures contain a nitro group?

  ![Nitro group](image)

• Mathematically, this is a subgraph isomorphism problem

• Requires way of representing query fragment(s)
Similarity search

- Looking for all the structures in a database that are highly similar to a given structure
- e.g. show me structures with a similarity greater than 0.7 to this molecule

- Requires a way of measuring similarity
- Solved using fingerprint representations and similarity coefficients
www.molinspiration.com/cgi-bin/search

molinspiration

Structure and Similarity Search

Draw your target molecule, choose search type from the menu below, and press the [Start Search] button. By problems with the structure input or display (your browser does not support Java?) go here. More information about the search technology is available here.

JME Editor courtesy of Peter Ertl, Novartis

JME help

Search type

Substructure

Start Search
Substructure search results

Substructure search, 4 hits found, 33727 molecules processed, search took 1.8 seconds.

Please contact Molinspiration to get more information, or to arrange evaluation of our database software.

New search  Modify target molecule  About substructure and similarity searches  Molinspiration home
Similarity search results

Similarity search, 25 most similar molecules shown, 33727 molecules processed, search took 2.6 seconds.
Representation of 3D chemical structures

[Chemical structure diagram with atom labels and coordinates table]
rCDK – R package for cheminformatics

- Gives access to CDK functionality in R
- CDK = Chemistry Development Kit [https://sourceforge.net/projects/cdk/](https://sourceforge.net/projects/cdk/)
- For demo see [http://dsdht.wikispaces.com/Using+chemistry+data+in+R](http://dsdht.wikispaces.com/Using+chemistry+data+in+R)

**Cheminformatics**

- 2D diagram editing and generation
- 3D geometry generation
- substructure search using exact structures and SMARTS-like queries
- QSAR descriptor calculation[16]
- fingerprint calculation, including the ECFP and FCFP fingerprints[17]
- force field calculations
- many chemical input/output formats, including SMILES, CML, and MDL formats
- structure generators[18]
- International Chemical Identifier support (via JNI-InChI)

**Bioinformatics**

- protein active site detection
- cognate ligand detection[19]
- metabolite identification[20]
- pathway databases
- 2D and 3D protein descriptors[21]

**General**

- Python wrapper (see Cinfony)
- Ruby wrapper
- active user community
What is a macromolecule?

- Any very large molecule (>1000 atoms)
- Usually made up of repeating building block molecules (amino acids, nucleic bases, etc) in a chain
- Polypeptides (amino acid building blocks)
- Proteins (amino acid building blocks)
- Nucleic acids (made up of bases)
- Polysaccharides (made up of sugars)
- We shall be focusing on polypeptides and proteins
Types of protein information

- Atomic (3D atom coordinates and bond information)
- Primary (Amino acid sequence)
- Secondary (Alpha helices, beta sheets, etc)
- Tertiary (3D folding of protein)
- Quaternary (dimers, protein families)
Atomic information

• 3D coordinates of all atoms in the protein
• Derived from X-ray crystallography or NMR Spectroscopy
Primary structure (Sequence)

- Lists Amino acids in order they appear in chain
- Uses three letter or one-letter abbreviations, e.g:

  Ser-Tyr-Ser-Met-Glu-His-Phe-Arg-Trp-Gly-Lys

  S Y S M E H F R W G K

- Essentially “1-dimensional” representation of the protein
- Can be stored on computer as a text string
Secondary structure

Certain groups of amino acids tend to form themselves into regular 3D shapes:

- **α-helix** – C=O and NH groups hydrogen bond to group 4 along in the chain, forming a coil shape: β-sheet, turn

- **β-sheet** – flat structure due to hydrogen-bonding between two or more chains
Secondary structure (2)

- Secondary structural features can be fairly well predicted from primary structure, or it can be inferred from atom coordinates.
- Primary sequence can be ‘tagged’ with secondary structure information.

- E.g.

  G A F T G E I S P G M I K D C G A T W V

  β β β β β β β  α α α α α α α α
Tertiary structure

• How the protein chain is folded in three dimensions
• Information mostly derived from atomic coordinate information
• Extremely difficult to predict from scratch using computational methods
• May be predicted by finding proteins with similar primary and secondary structures that have known coordinates (*homology modeling*, *threading*).
Tertiary structure example (HIV)
Big Data in the public domain

- There is now an incredibly **rich resource of public information** relating compounds, targets, genes, pathways, and diseases. Just for starters there is in the public domain information on:
  - 69 million compounds and 449,392 bioassays (PubChem)
  - 59 million compound bioactivities (PubChem Bioassay)
  - 4,763 drugs (DrugBank)
  - 9 million protein sequences (SwissProt) and 58,000 3D structures (PDB)
  - 14 million human nucleotide sequences (EMBL)
  - 19 million life science publications - 800,000 new each year (PubMed)
  - Multitude of other sets (drugs, toxicogenomics, chemogenomics, metagenomics ...)

- Even more important are the **relationships between these entities**. For example a chemical compound can be linked to a gene or a protein target in a multitude of ways:
  - Biological assay with percent inhibition, IC50, etc (e.g. ChEMBL, PubChem)
  - Crystal structure of ligand/protein complex
  - Co-occurrence in a paper abstract
  - Computational experiment (docking, predictive model)
  - Statistical relationship
  - System association (e.g. involved in same pathways cellular processes)
Chem2Bio2RDF.org

- NCI Human Tumor Cell Lines Data
- PubChem Compound Database
- PubChem Bioassay Database
- PubChem Descriptions of all PubChem bioassays
- Pub3D: A similarity-searchable database of minimized 3D structures for PubChem compounds
- Drugbank
- MRTD: An implementation of the Maximum Recommended Therapeutic Dose set
- Medline: IDs of papers indexed in Medline, with SMILES of chemical structures
- ChEMBL chemogenomics database
- KEGG Ligand pathway database
- Chem2Bio2RDF
- Comparative Toxicogenomics Database
- PhenoPred Data

31m chemical structures
59m bioactivity data points
19m publications
~5,000 drugs

BMC Bioinformatics, 2010, 11, 255; chem2bio2rdf.org
Finding association paths with DFS searches

• Given any two entities in the Chem2Bio2RDF network (e.g. compound & disease, drug & side-effect), find all possible shortest paths of \( \leq n \) edges between them.

• Can be weighted by literature support using BioLDA algorithm.

• Can be constrained by entity or relational type (e.g. from ontology), data source, number of paths, or as a diverse subset of paths or graphs.

• Incorporated in interactive Flash-based tool (available at http://djwild.info).

• For more information, see *PLoS One*, in review.
Association search: Rosiglitazone and M.I.
Semantic Link Association Prediction (SLAP)

- Predicts a *probability of association* of a compound and a target based on the network paths between them that involve drugs, targets, pathways, diseases, tissues, GO terms, chemical ontologies, substructure and drug side-effects
- It can be primarily considered as a “missing link prediction”
- Data source is a subset of the Chem2Bio2RDF network including 250,000 compounds with known bioactivities and the targets known to be associated with these drugs
- *Raw Score* is a measure of the significance of paths between a compound and target, based on topology and semantics of the path nodes and edges. Raw scores are normally distributed within a *path pattern*
- *Association Score* is a sum of z-scores of raw scores relative to a distribution of random pair scores for different paths and path patterns. Association scores form a normal distribution
- *Association Significance* is a significance p-value of an association score based on the normal distribution of association scores.
Example: Troglitazone and PPARG

Troglitazone

- Chemical ontology
- bind
- hypoglycemic drug
- bind
- Rosiglitazone
- bind

Eicosapentaenoic Acid

- Chemical ontology
- bind
- PPARA
- bind
- bind
- bind
- Pioglitazone
- bind

PPARG

- Response to nutrient
- GO
- pathway

ACSL4

- PPAR signaling pathway
- GO
- pathway

=> missing link predicted
ChemoHub – fusing prediction methods

• Work in progress as data fusion of multiple ways of predicting biological effects and gene associations of a compound

• Currently allows calculation of several quantitative measures of compound-gene association, including SLAP, SEA (Nature 462, 175-181), Bayesian analysis and literature co-occurrence.

• Provides network view of association using Cytoscape plug-in. Nodes and edges can be clicked to get to raw data
Chemical & Biological Literature Extraction

Covering 1865-2009
18,502,916 PubMed/Medline literature records!

Table 1. Statistics of the bio-terms extraction.

<table>
<thead>
<tr>
<th>Bio-Terms</th>
<th># of unique terms</th>
<th># of term-citation pairs</th>
<th># of unique citations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compound</td>
<td>56,383</td>
<td>11,775,891</td>
<td>5,856,084</td>
</tr>
<tr>
<td>Drug</td>
<td>2,820</td>
<td>5,624,529</td>
<td>3,427,067</td>
</tr>
<tr>
<td>Gene</td>
<td>13,022</td>
<td>5,252,844</td>
<td>3,735,517</td>
</tr>
<tr>
<td>Disease</td>
<td>3,848</td>
<td>12,612,636</td>
<td>7,066,084</td>
</tr>
<tr>
<td>Side Effect</td>
<td>1,363</td>
<td>10,489,676</td>
<td>6,310,741</td>
</tr>
<tr>
<td>Pathway</td>
<td>180</td>
<td>916,754</td>
<td>838,090</td>
</tr>
</tbody>
</table>

Wang HJ 2010

PLoS One, 6 (3), e17243
BioLDA Topic Model of PubMed Literature

- Latent Dirichlet Association (LDA) identifies “latent topics” by word association: a kind of fuzzy clustering. Each word can have associations with multiple topics, and has a varying degree of strength.
- Term-topic edges are labeled with probability (i.e. strength of a relationship to a topic). Term-term edges labeled with KL-divergence (measure of distance).
- We considered BioTerms rather than free text, and applied to 336,899 MedLine abstracts on 50 topics published in 2009.
- Based on work done by Jie Tang on social networks (see www.arnetminer.com).
- More information can be found in *PLoS One*, 2011, 6 (3), e17243.
Example: Topic 10

| Topic Assignment | P(z|b) |
|------------------|-------|
| Venlafaxine      | 0.9676|
| HTR1A            | 0.9296|
| HTR2A            | 0.5871|
| Depressive Disorder | 0.9981|
| Anxiety Disorder | 0.9445|
| Obsessive-Compulsive Disorder | 0.9782|

\[
sKL(\text{"Venlafaxine","HTR1A"}) = 0.34
\]
\[
sKL(\text{"Venlafaxine","HTR2A"}) = 4.0
\]
Two worlds of medical data – never the twain shall meet?

Molecular Data

Chemistry, Biology, Toxicology
Chemical compounds / Drugs
Proteins and Genes
Bioassay
Gene expression / MicroArray

Patient Data

Physicians, Clinical Trials, Patients
Observed side effects & adverse events
Observational clinical studies
Electronic medical records
Epidemiology & demographic
Phenotypic data
Web & social media
<table>
<thead>
<tr>
<th>Rank</th>
<th>Status</th>
<th>Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Completed</td>
<td>RECORD: Rosiglitazone Evaluated for Cardiac Outcomes and Regulation of Glycaemia in Diabetes</td>
</tr>
<tr>
<td></td>
<td>Has Results</td>
<td><strong>Condition:</strong> Diabetes Mellitus, Type 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Interventions:</strong> Drug: Rosiglitazone; Drug: Sulfonylurea; Drug: Metformin</td>
</tr>
<tr>
<td>2</td>
<td>Unknown</td>
<td>Defining the Role of Insulin Resistance in 'idiopathic' Dilated Cardiomyopathy</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Condition:</strong> Dilated Cardiomyopathy</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Intervention:</strong> Drug: Rosiglitazone therapy</td>
</tr>
<tr>
<td>3</td>
<td>Completed</td>
<td>Role of Rosiglitazone on Pre-Diabetes Mellitus and Coronary Artery Disease</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Conditions:</strong> Prediabetes; Coronary Artery Disease; Insulin Resistance; Glucose Intolerance</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Interventions:</strong> Drug: placebo tablet; Drug: rosiglitazone (4 mg/day)</td>
</tr>
<tr>
<td>4</td>
<td>Completed</td>
<td>The DREAM (Diabetes Reduction Assessment With Ramipril and Rosiglitazone Medication) Trial</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Conditions:</strong> Impaired Glucose Tolerance; Cardiovascular Disease; Glucose Metabolism Disorders</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Interventions:</strong> Drug: Ramipril; Drug: Rosiglitazone</td>
</tr>
</tbody>
</table>
## Rosiglitazone maleate

### Side effects and indications

Whenever possible, frequency information about the side effects was extracted from the labels. Aggregated frequency information for the drug and, if available, from the placebo labels is shown. Click on shaded boxes to be taken to mentions of the side effect on the label. In some cases, the side effect cannot be highlighted due to conversion problems of the data extracted from the indications and usage sections of the labels.

### Show MedDRA Preferred Terms

<table>
<thead>
<tr>
<th>Side effect</th>
<th>Data for drug</th>
<th>Placebo Labels (show all 15)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Respiratory tract infection</strong></td>
<td>postmarketing, 7.3% - 9.9%</td>
<td>8.7%</td>
</tr>
<tr>
<td><strong>Upper respiratory tract infection</strong></td>
<td>postmarketing, 7.3% - 9.9%</td>
<td>8.7%</td>
</tr>
<tr>
<td><strong>Hyperglycaemia</strong></td>
<td>postmarketing, 3.9% - 8.1%</td>
<td>5.7%</td>
</tr>
<tr>
<td><strong>Headache</strong></td>
<td>postmarketing, 5.4% - 5.9%</td>
<td>5%</td>
</tr>
<tr>
<td><strong>Hypoglycaemia</strong></td>
<td>postmarketing, 0.6% - 5.9%</td>
<td>0.2%</td>
</tr>
<tr>
<td><strong>Back pain</strong></td>
<td>postmarketing, 4% - 5%</td>
<td>3.8%</td>
</tr>
<tr>
<td><strong>Pain</strong></td>
<td>postmarketing, 4% - 5%</td>
<td>3.8%</td>
</tr>
<tr>
<td><strong>Fatigue</strong></td>
<td>1.9% - 3.6%</td>
<td>5%</td>
</tr>
<tr>
<td><strong>Sinusitis</strong></td>
<td>postmarketing, 3% - 3.2%</td>
<td>4.5%</td>
</tr>
<tr>
<td><strong>Diarrhoea</strong></td>
<td>postmarketing, 2.3% - 3%</td>
<td>3.3%</td>
</tr>
<tr>
<td><strong>Abdominal pain</strong></td>
<td>postmarketing</td>
<td></td>
</tr>
<tr>
<td><strong>Anaphylactic shock</strong></td>
<td>postmarketing</td>
<td></td>
</tr>
<tr>
<td><strong>Anaemia</strong></td>
<td>postmarketing</td>
<td></td>
</tr>
</tbody>
</table>
FDA Adverse Event Reporting System (FAERS) (formerly AERS)

What is FAERS?

The FDA Adverse Event Reporting System (FAERS) is one of the largest databases of its kind in the world. The FAERS database is a program for collecting spontaneous reports of adverse drug events, deaths, and serious injuries occurring to human beings in the United States. While the FAERS database contains information from other countries, the reporting system is used only in the United States.

How Does FAERS Work?

The FAERS database is used to identify trends in adverse drug events, deaths, and serious injuries occurred to human beings in the United States. The FAERS database is used to identify trends in adverse drug events, deaths, and serious injuries occurred to human beings in the United States. The FAERS database is used to identify trends in adverse drug events, deaths, and serious injuries occurred to human beings in the United States.

Who Reports to FAERS?

The FAERS database is used to identify trends in adverse drug events, deaths, and serious injuries occurred to human beings in the United States. The FAERS database is used to identify trends in adverse drug events, deaths, and serious injuries occurred to human beings in the United States. The FAERS database is used to identify trends in adverse drug events, deaths, and serious injuries occurred to human beings in the United States.

Resources for You

- MedWatch: The FDA Safety Information and Adverse Event Reporting Program
- Drug Safety and Availability

![Graph showing the number of reports by year and type of report (Direct, Expedited, Non-Expedited).]
Epidemiology – e.g. CDC WONDER

- **What is WONDER?**
- **Frequently Asked Questions**
- **Data Use Restrictions**
- **Data Collections**
- **Citations**
- **Republishing WONDER Data**
- **What's New?**

**WONDER Systems**

- **Topics**
- **A-Z Index**

**WONDER Online Databases**

- AIDS Public Use Data
- Births
- Cancer Statistics

**Environment**

- Daily Air Temperatures & Heat Index
- Daily Land Surface Temperatures
- Daily Fine Particulate Matter
- Daily Sunlight
- Daily Precipitation

**Mortality**

- Underlying Cause of Death
  - Detailed Mortality
  - Compressed Mortality
  - Multiple cause of death (Detailed Mortality)
  - Infant Deaths (Linked Birth/Infant Death Records)
- Online Tuberculosis Information System

**Population**

- Bridged-Race Population (from NCHS)
- Population (from Census)
- Sexually Transmitted Disease Morbidity

**Reports and References**

- Prevention Guidelines (Archive)
- Scientific Data and Documentation (Archive)

**Other Query Systems**

- Healthy People 2010
- MMWR Morbidity Tables
- MMWR Mortality Tables
Social media extraction

The person uses Claritin

Analysis of 4,900 pieces of content

Relevance:
- Yes: 4,663
- No: 12
- Non-English: 225

Sentiment:
- Very negative: 2,327
- Neutral: 0
- Positive: 0
- Very positive: 0

Gender of author:
- Female: 2,555
- Male: 1,556
- Unknown: 552

Timeline:

Relevance to:
- Dizziness: Yes - 11, No - 4,652
- Convulsions: Yes - 4,663, No
- Heart palpitations: Yes - 4,658, No
- Shortness of breath: Yes - 4,659, No
- Headaches: Yes - 4,655, No
- Drug effect decreased: Yes - 4,597, No
- Allergies worse after: Yes - 132, No - 4,531
- Bad interaction between: Yes - 40, No
- Nausea (made the person feel...): Yes - 4,659, No
- Caused insomnia (the person...): Yes - 26, No
Electronic Medical Records

[Image of a computer interface showing a patient's medical records, including sections for problem list, visit diagnosis, and historical diagnosis.]

- **Patient Chart**
  - **Demo_Father**
  - **5485**
  - **05-Mar-1955**
  - **50**
  - **M**

- **Chart Review**
  - **28-Jul-2005 10:15**
  - **USER_DEMO**

- **Problem List**
  - **ID**: SOUC3
  - **Provider Narrative**: TYPE 2 DIABETES MELLITUS
  - **Status**: Active
  - **Entered**: 03/11/2000
  - **Onset**: 03/11/2000
  - **Notes**: In spite of regular exercise, I'm putting client on medication.

- **ICD Pick-Lists**
  - Administrative
  - Medicine Pick List
  - Obstyn Pick List
  - Optometry
  - Pediatrics Pick List

- **Historical Diagnosis**

- **Visit Diagnosis**
  - **Provider Narrative**: TYPE 2 DIABETES MELLITUS
  - **ICD**: 250.00
  - **ICD Name**: DM UNCOMP/II/NIDDMNS
  - **Priority**: Primary

- **Chief Complaint**: I broke my ankle (cases demo)
  - **Vitals**: WT: 200 (91 kg), HT: 65 (156 cm), TMP: 95.8 (98.7), BP: 120/80, PU: 72, RS: 16, PA: 7, C/O: 5, BMI: 33.3
  - **Immunizations**: DTAP

- **Notifications**
  - **Cover Sheet**
  - **Triage**
  - **Wellness**
  - **Notes**
  - **Services**
  - **Prob/POV**
  - **Orders**
  - **Medications**
  - **Labs**
  - **D/C Summ**
  - **Reports**
  - **Consults**

- **Interface Details**
  - **User Demo**
  - **Demo.CLINFORMTICS.COM**
  - **Demo Hospital**
  - **15-Aug-2005 16:59**